

AeDES: A Next-Generation Monitoring and Forecasting System for Environmental Suitability of *Aedes*-borne Disease Transmission

Á.G. Muñoz^{1,*}, X. Chourio¹, Ana Riviere-Cinamond², M. Diuk-Wasser³, P. Kache³, E. Mordecai⁴, L. Harrington⁵, and M.C. Thomson⁶

¹International Research Institute for Climate and Society (IRI). The Earth Institute at Columbia University. Palisades, NY, 10964. USA.

²Pan-American Health Organization (PAHO), World Health Organization (WHO). Washington, DC, USA.

³Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, 10027, USA.

⁴Biology Department, Stanford University, Stanford, CA, postcode, USA.

⁵Department of Entomology, Cornell University, Ithaca, NY, 14853, USA.

⁶Wellcome Trust, London, NW1 2BE, UK.

*Corresponding author: agmunoz@iri.columbia.edu

ABSTRACT

Aedes-borne diseases, such as dengue and chikungunya, are responsible for more than 50 million infections worldwide every year, with an overall increase of 30-fold in the last 50 years, mainly due to city population growth, more frequent travels and ecological changes. In the United States of America, the vast majority of *Aedes*-borne infections is imported from endemic regions by travelers, who can become new sources of mosquito infection upon their return home if the exposed population is susceptible to the disease, and if suitable environmental conditions for the mosquitoes and the virus are present. Since the susceptibility of the human population can be determined via periodic monitoring campaigns, the environmental suitability for the presence of mosquitoes and viruses becomes one of the most important pieces of information for decision makers in the health sector. We present a next-generation monitoring and forecasting system for *Aedes*-borne diseases' environmental suitability (AeDES) of transmission in the conterminous United States and transboundary regions, using calibrated ento-epidemiological models, climate models and temperature observations. After showing that the seasonal predictive (discrimination) skill of AeDES is ~1.4-1.8 times larger than the baseline, we illustrate how a combination of tailored deterministic and probabilistic forecasts can inform key prevention and control strategies.

We live in an increasingly interconnected world. The rapidly increasing movement of people, pathogens, vectors, livestock, food, goods, and capital across borders creates both economic opportunities and health risks¹. Epidemics, just like climate, do not respect national borders and can threaten human health and social stability. Since the Millennium, the appearance of the SARS coronavirus in 2003, the avian influenza (H1N1) in 2009, the Ebola virus in West Africa (2014-2016), the Zika virus in the Americas (2015) and the novel coronavirus identified in late December 2019 in Wuhan, China and still ongoing (2020), amongst others, has demonstrated the speed at which emerging infectious diseases can spread with devastating effects².

Many infectious diseases are climate-sensitive; climate acting as an important driver of spatial and seasonal patterns of infections, year-to-year variations in incidence (including epidemics), and longer-term shifts in populations at risk³. Climate impacts both the virus and the vector. Evidence to date shows that arboviruses of global public health importance, including Zika, dengue, yellow fever, chikungunya, and Rift Valley Fever, have mosquitoes as part of their epidemiological cycles.

Some *Aedes*-borne diseases have experienced an overall increase of 30-fold in the last 50 years, causing more than 50 million infections worldwide every year⁴. In the United States of America, the vast majority of *Aedes*-borne infections is imported from endemic and often neighboring regions -like the Caribbean, Central and South America- by travelers who become potential new sources of transmission. Autochthonous transmission in the continental USA has been already observed for chikungunya virus (2013) and Zika (2017), and risks are likely to increase with anthropogenic global warming.

For autochthonous transmission to occur, the population needs to be susceptible to the disease, but there must also be suitable environmental conditions (e.g., suitable temperatures) for both the mosquitoes and the virus. Since the susceptibility can be periodically monitored via targeted campaigns, the environmental suitability for presence of mosquitoes and viruses is one of the most important pieces of information for decision makers in the health sector. Moreover, the transmission rates or the number of cases are generally more difficult to forecast than environmental suitability, due to their link to a larger number of (often entangled and more complex) predictors, involving human behavior and socio-economic conditions.

A generalized approach to model *Aedes*-borne pathogens is needed because multiple *Aedes* can serve as vectors of dengue, Zika and chikungunya. Although *Aedes aegypti* is the most common vector, *Aedes albopictus* (otherwise known as the Asian tiger mosquito) has been identified as another important additional vector because of its vector competence for several arboviruses and recent rapid spread⁵. Both vectors pose a potent threat to global health security given their ability to transmit a wide variety of emerging and re-emerging arboviruses for which there are no vaccines. *Aedes aegypti* and *Aedes albopictus* are ubiquitous in large regions of the Americas and the Caribbean.

Historical, current and forecast climate information can be combined with disease models to improve climate-sensitive health planning and targeting of resources. For infectious disease models, the goal has frequently been to explore different interventions scenarios in order to help set priorities for policy makers⁶. However, in recent years there is increasing interest in using models for real-time forecasting^{7–9}, although there remains a significant gap in the operational readiness of the numerous forecasting systems presented in the literature¹⁰. Stochastic models possess inherent randomness and are widely used in climate science as well as in disease modelling to build probabilistic forecasts¹¹, as they provide a more reliable assessment of the range of likely outcomes. However, probabilistic models are sometimes harder for decision-makers to interpret, and tend to be rejected in favour of simpler, deterministic, but over-confident, models. An approach that takes full advantage of both deterministic and probabilistic forecasts is presented and discussed in the following pages.

Although the historical (average) seasonal behavior –and similar statistics– of these diseases is useful¹², we consider it not enough for decision-making, as inter-annual variability (e.g., related to El Niño–Southern Oscillation) tends to play an important role in the actual observed variations of *Aedes*-borne diseases, enhancing or reducing the associated risk^{8,13–16}. Hence, a formal forecast system and its associated skill assessment is required and –to the best of our knowledge– is still nonexistent for the continental United States and its transboundary regions.

Here, we describe the AeDES (*Aedes*-borne diseases' environmental suitability) system, a new pattern-based calibrated, multi-model ensemble of climate-driven *Aedes*-borne disease models for North America, Central America, northern south America and the Caribbean, based on prior work undertaken in collaboration with the Pan American Health Organization (PAHO)/World Health Organization (WHO)^{8,16,17}. We built AeDES using the same general approach for both the monitoring and forecasting sub-systems, which in addition to supporting surveillance operations, simplifies the forecast verification process. We discuss the use of AeDES to inform concrete prevention and control strategies, using the recent Zika epidemic as an example.

Results

AeDES uses multiple ento-epidemiological models to produce estimations of environmental suitability for transmission of *Aedes*-borne diseases, quantified via the basic reproduction number, $\mathcal{R}_0$ (red box in Figure 1). We used the basic reproduction number to assess the environmental suitability of transmission of *Aedes*-borne diseases because (a) it is one of the operational outbreak indices used by WHO and several other decision-making institutes and health practitioners^{18,19}, and (b) it has an intuitive interpretation in terms of the number of secondary human cases one case generates on average over the course of its infectious period (assuming a completely susceptible population)²⁰; hence, values smaller than one indicate that environmental conditions are not suitable for disease propagation.

Formally speaking, $\mathcal{R}_0$ is an environmental suitability (or potential) for transmission, and not a transmission risk index itself; the latter depends on more complex interactions and the definition of the involved vulnerability. $\mathcal{R}_0$ works both as a suitability monitoring index –when computed using observed variables, or when estimated by an authoritative organization such as PAHO or the Center for Disease Control (CDC)—, and as a forecast index –when using actual climate forecasts of the variables required for its computation.

$\mathcal{R}_0$ models require a set of ento-epidemiological parameters (green box, in Figure 1) and environmental information, either actual observations if we focus on the monitoring sub-system, or forecasts if we focus on the prediction sub-system (see blue box in Figure 1). Typically, $\mathcal{R}_0$ models require near-surface (2 meter) temperatures, but other environmental variables are also involved, like rainfall or even humidity. Here, we use four $\mathcal{R}_0$ models already described in the literature: the Caminade et al²¹, Wesolowski et al²², Liu-Helmersson et al²³ and Mordecai et al²⁴ models. For details see the Methods section. We use multiple $\mathcal{R}_0$ models to be able to better assess uncertainties, and we calibrate each of the models independently before creating the multi-model ensemble to minimize systematic errors.

Monitoring sub-system

The AeDES monitoring sub-system offers maps showing the spatial distribution of environmental suitability over the region of study for the 1948–present period, at a monthly timescale. It also includes additional information to provide context to the user (Figure 2). These fields were included in the AeDES Maproom (<https://aedes.iri.columbia.edu>) after consultation with decision-makers at PAHO.

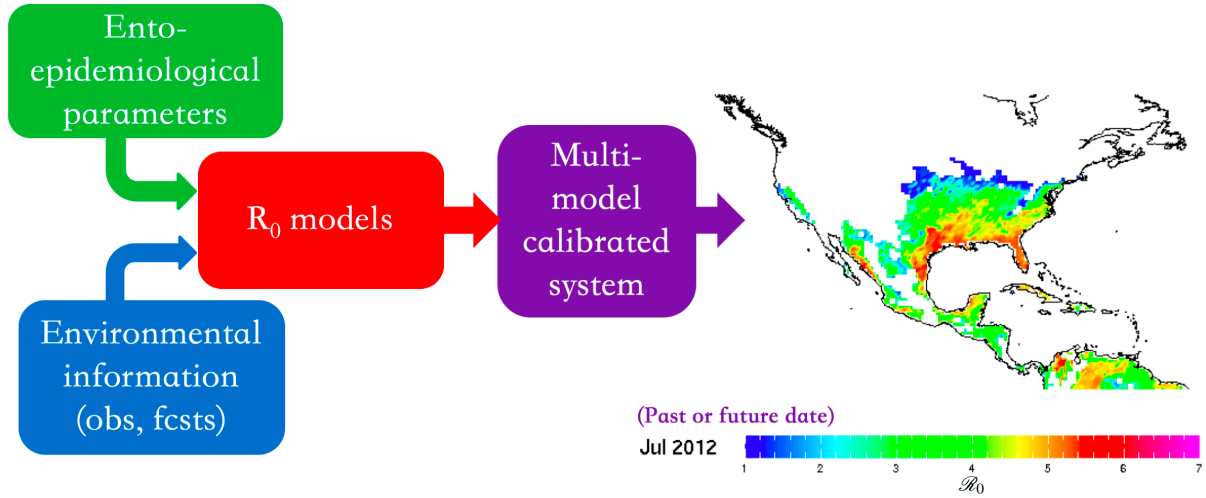


Figure 1. Monitoring and forecast system schematics. Ento-epidemiological and environmental information (obs: climate observations, fcsts: climate forecasts) are used to force four $\mathcal{R}_0$ models. Each model is independently calibrated using a pattern-based post-processing approach before being combined.

To produce the environmental suitability maps (e.g., Figure 2a), each one of the four $\mathcal{R}_0$ models was run from 1948 to present, forced by GHCN-CAMS temperature data²⁵ (~56 km resolution) and ento-epidemiological parameters (see Methods), and then combined. The monitor sub-system is automatically updated in the AeDES Maproom around the 8th day of each month. These maps are useful to know the recent behavior of environmental suitability, or to conduct comparisons with respect to particular years. Trends and variability analysis, or the extension of the northern border of environmental suitability can be easily too computed with this new dataset.

The additional information, such as population density (Figure 2b), and social vulnerability (Figure 2c) is offered to the user to assess potential risk of transmission. Once a location is selected, the seasonality of $\mathcal{R}_0$, accumulated rainfall, minimum, average and maximum temperatures, and frequency of rainy days (Figures 2d-g) is provided. Our team is working on adding fields such as human mobility and connectivity, which local experts in the northeast of the US have suggested as also useful to analyze potential disease transmission.

Forecast sub-system calibration and evaluation

As indicated earlier, the forecast sub-system employs the output of state-of-the-art climate models and the same $\mathcal{R}_0$ models used by the monitoring system. Models, nonetheless, require statistical post-processing to help correct for biases with respect to the monitored $\mathcal{R}_0$ values. Following Muñoz et al⁸, a pattern-based Model Output Statistics (MOS) approach using principal component regression (PCR) was applied to the raw $\mathcal{R}_0$ models output. Since the $\mathcal{R}_0$ models are the same (using the same ento-epidemiological parameters) the calibration takes care of climate-related model biases only.

A skill assessment was conducted for each calibrated $\mathcal{R}_0$ model and the final multi- $\mathcal{R}_0$ model ensemble (i.e., the AeDES model), focusing on discrimination as an actual measure of the value of a forecast system²⁶. Although correlations between forecasts and observation are often used to assess skill, they only provide information of how in phase or not the forecasts

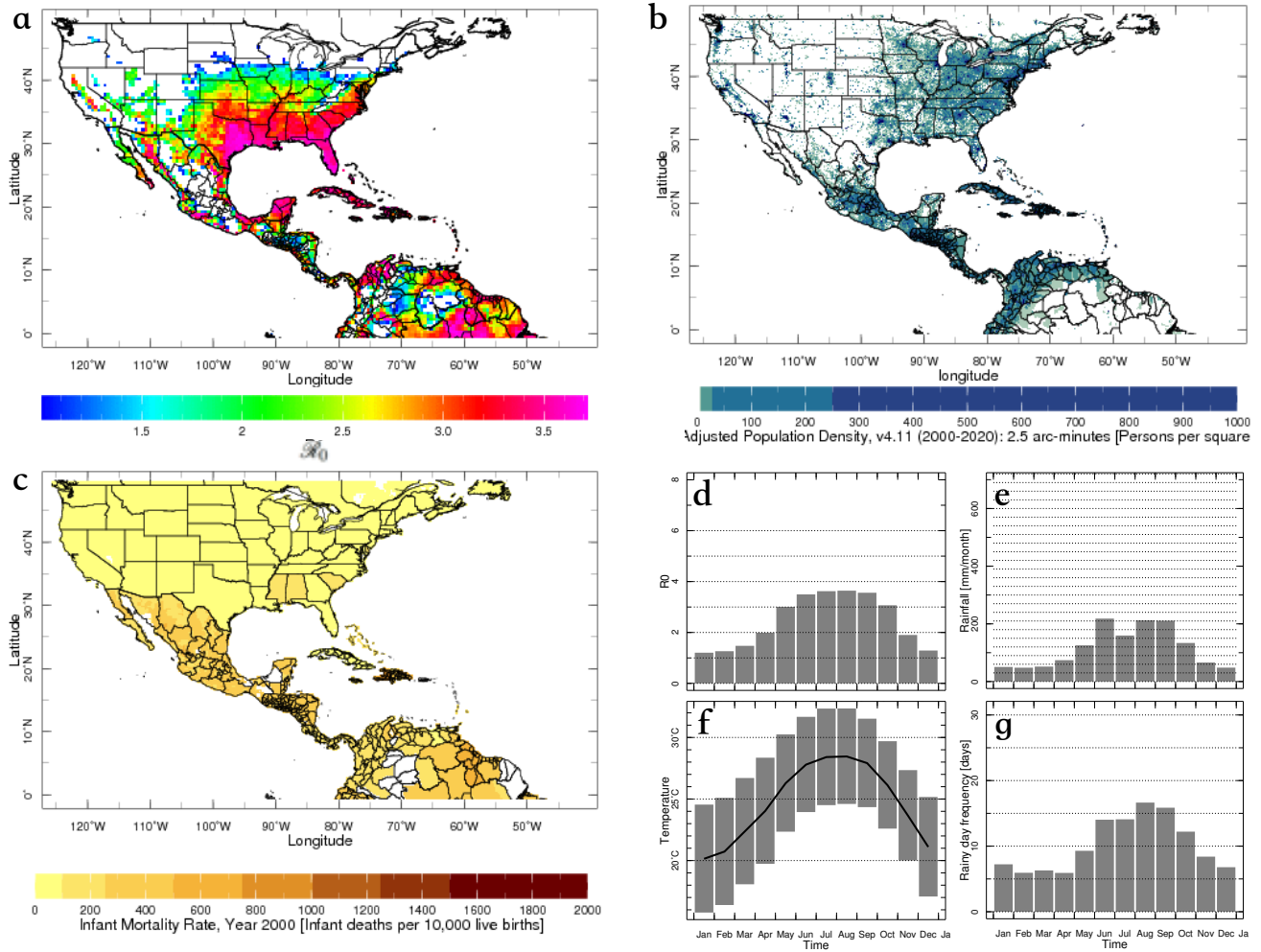


Figure 2. Example of fields available in the *AeDES* monitor system: (a) Basic reproduction number for July 2019 (only locations with $\mathcal{R}_0 > 1$, suitable for transmission, are plotted); (b) population density (persons per square kilometer; estimated for 2020); (c) infant mortality (infant deaths per 10,000 live births); and seasonality of (d) $\mathcal{R}_0$, (e) accumulated rainfall, (f) maximum, average and minimum temperature and (g) frequency of rainy days for Miami, FL (any other location in the map can be plotted). Only values corresponding to suitable conditions for transmission ($\mathcal{R}_0 > 1$) are plotted.

are with respect to observations. The metric selected for skill assessment was the two-alternative forced choice, or 2AFC, which "measures the proportion of all available pairs of observations of differing category whose probability forecasts are discriminated in the correct direction"²⁶. In other words, when terciles (above-normal, normal, below normal conditions) are used, the 2AFC measures how well the system distinguishes between the different categories; a system with poor discrimination is of no practical and economical value for decision-makers. Furthermore, 2AFC has an intuitive interpretation as an indication of how often the forecasts are correct²⁶.

AeDES's predictive skill (as measured by 2AFC) is well above that of the reference long-term average (corresponding to 2AFC=50%), with values ~ 1.4 - 1.8 times larger than that baseline basically everywhere in the region under study. Skillful regions extend farther north during the boreal summer (Jun-Aug, or JJA) due to more suitable areas for the vectors because of higher seasonal temperatures (see Figures 3a,c). Also, as expected, *AeDES* exhibits skill improvement compared to any of the models involved in its ensemble (Figure 3), which show comparable skill distributions among themselves, both in space and time. *AeDES* tends to outperform the individual models everywhere, but especially in the Caribbean (e.g., Cuba, Jamaica, Haiti and Dominican Republic) and in a lower degree in the United States Great Plains, southern Mexico, Colombia's Orinoquia and the northern Amazon in Brazil (Figure 3); it also outperforms its predecessor model for Latin America and the Caribbean, described by Muñoz et al.⁸, especially in summer in western Colombia, and in winter in most of Central America and the Yucatan Peninsula (cfr. Figure 4 in Muñoz et al.⁸).

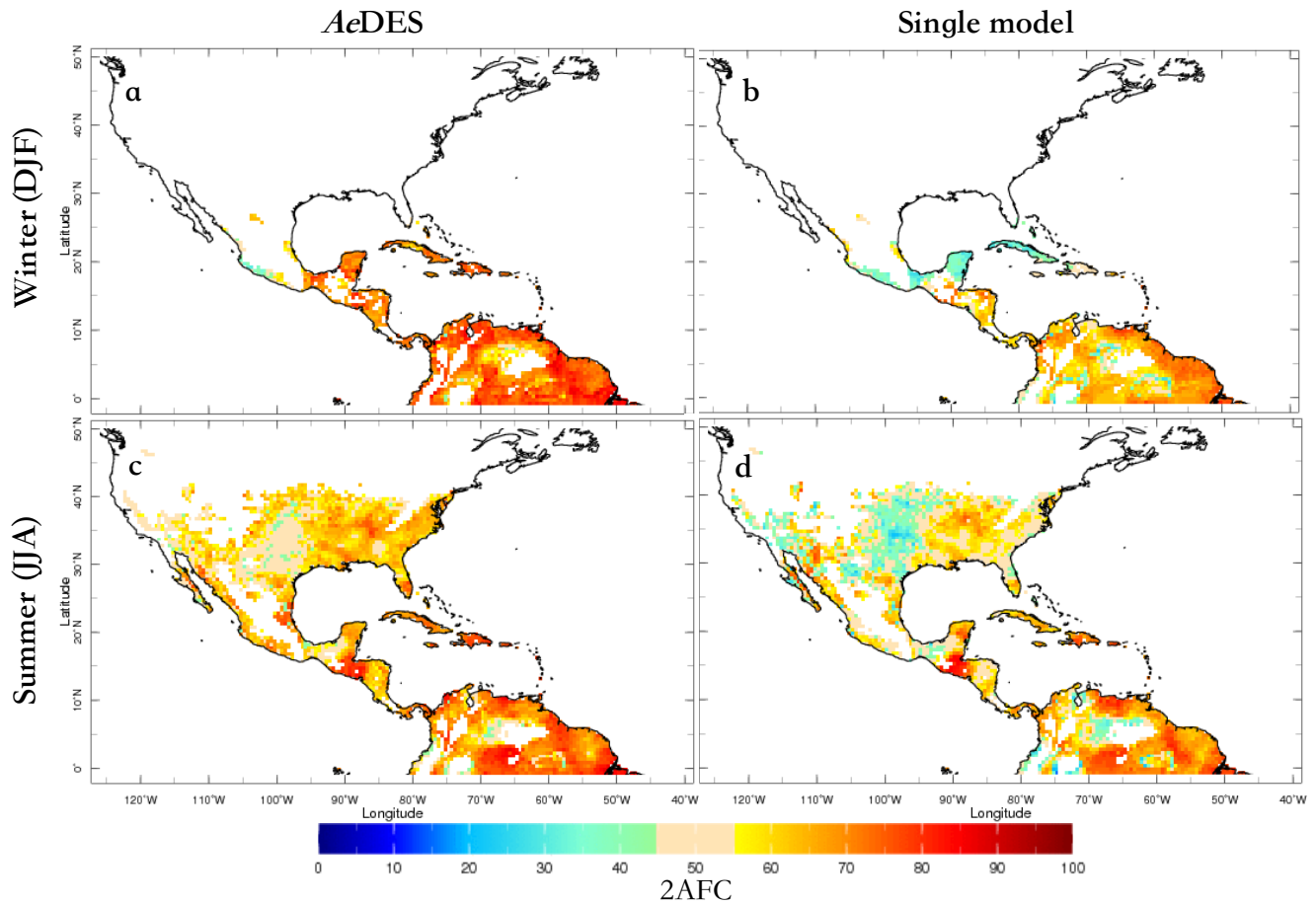


Figure 3. Cross-validated skill assessment (using 2AFC) between the AeDES multi- $\mathcal{R}_0$ model system (panels a,c) and the Caminade et al²¹ $\mathcal{R}_0$ model, for the boreal winter (a,b; DJF: Dec-Feb) and summer (c,d; JJA: Jun-Aug) seasons. Values above (below) 50% indicate better (worse) discrimination than long-term averages; only values corresponding to suitable conditions for transmission ($\mathcal{R}_0 > 1$) are plotted. Skill of the other involved models is similar to the Caminade et al²¹ one (see Data and Codes Availability).

Predictive skill of the AeDES system is especially high (2AFC $\sim 70\%-90\%$) in most locations of Central America, the Caribbean and northern South America in boreal winter (Dec-Feb, or DJF), with "skill hotspots" in both boreal summer and winter in Guatemala, Honduras, El Salvador, Cuba, Haiti and Dominican Republic, Jamaica, Puerto Rico and some island nations in the Lesser Antilles (unfortunately the observational dataset used for calibration does not cover all of these island nations).

Regarding North America, the Yucatan Peninsula is one of the locations with highest skill, especially in DJF, a peak season for tourism, and thus increased human mobility. In summer, almost the entire Pacific coast of Mexico exhibits 2AFC values above 65%. Overall, predictive skill over the United States in summer tends to be higher in the eastern half of the country than in the western half (where orographic temperatures naturally tends to control vector proliferation in large regions), and ranges between 50% and 90% along the United States-Mexico border and the states along the Gulf of Mexico's shoreline. Forecast discrimination skill for southern Florida is also high in summer (values $\sim 90\%$, see Figure 3c). In northern South America, the Caribbean coast of Colombia, and northern regions of Venezuela, Guyana, Suriname, French Guyana and northeastern Brazil exhibit very high skill both in summer and winter.

Hence, predictive discrimination skill of AeDES is in general high, and decision makers geographically interested in the hotspots mentioned above can take special advantage of the enhanced skill of the system in these regions to improve their response times on key prevention and control strategies, at least a month ahead of the target season.

Discussion

The risk of *Aedes*-borne disease transmission is in general very difficult to estimate, in part due to the complexities of accurately assessing the actual risk in terms of hazards and vulnerabilities impacting the target population. The general approach should successfully integrate the interactions between humans, virus, vectors and the environment, making it a very complex system to forecast, not to mention the fact that a lot of those interactions are not yet well understood. An alternative is to identify a predictand (variable to monitor and predict) that (a) enables decision-makers to take timely, "no-regrets" actions, (b) is verifiable (can be easily obtained from the information available or the health surveillance systems in place), and (c) can be skillfully predicted for the region and timescales of interest. The information provided to decision-makers does not need to be perfect, it needs to be reliable enough to make the best decisions.

Typical choices of predictands in the case of interest are number of positive cases and incidence. Although these options generally satisfy the criteria (a) and (b) mentioned above, skill tends to be a barrier to make the best decisions in a timely manner. Low predictive capacity for these predictands is related to different reasons, but often can be traced back to the fact that they depend on a variety of complex factors -e.g., socio-economic conditions, human behavior, human mobility, etc.-, some of which are (still) largely unpredictable. Previous work have argued⁸ that a potential alternative is to focus on environmental suitability for transmission, since variables like temperature, relative humidity, vegetation cover and, depending on where, rainfall, are skillfully predictable at timescales decision-makers are interested in. In this sense, climate imparts predictability to the *Aedes*-borne diseases transmission problem if a predictand like $\mathcal{R}_0$ is used as a proxy for transmission risk, even when clearly it is not representing the complete risk picture: additional information on the presence of the vector(s), the population exposed to the disease, and circulation of the virus is also needed. Recent work by Monaghan et al.²⁷ is using a similar approach to the one presented here to address the vector presence/absence component of the problem, and certainly both systems could be combined to provide additional information for decision makers in the health sector.

AeDES, uncertainties and decision-making

A large amount of work in the related scientific literature has been focused on developing or improving different $\mathcal{R}_0$ models (see Van den Driessche²⁸ and references therein), but few efforts have addressed real-time $\mathcal{R}_0$ seasonal forecasts, and no such operational system -to the best of our knowledge- existed until now for *Aedes spp.* in North America, Central America, northern South America and the Caribbean basin. Furthermore, to better assess uncertainty in AeDES, the approach followed here involves the use of not one but multiple ento-epidemiological models, forced by state-of-the-art seasonal climate models from the National Oceanic and Atmospheric Administration (NOAA) North American Multi-Model Ensemble project (Kirtman et al., 2014).

There is consensus on the need of including uncertainty information on any forecast that is produced²⁹. One way of providing that information is to add confidence limits if the forecasts are deterministic (actual values of $\mathcal{R}_0$ in our case). For an example, see Figure 4, sketching the expected value for the summer of 2016 (Figure 4a), and the expected standard deviation (or uncertainty, Figure 4b); for reference, the monitored (or "observed") values for the same summer are presented in Figure 4c).

Another way to provide information about the forecast uncertainty is via the use of probabilities to indicate how confident (or not) the system is that a certain outcome -say, above normal $\mathcal{R}_0$ values- will occur during the next season. An example of a tercile-based $\mathcal{R}_0$ probabilistic seasonal prediction, again for the summer of 2016, is presented in Figure 5a, where probabilities of below-normal, normal or above-normal $\mathcal{R}_0$ values correspond to red, green and blue color shades, respectively. Although this is a very useful approach, and tercile-based probabilistic forecasts have been used for more than two decades now, decision makers often require information beyond the usual three categories described above. Using the entire probability density function (see Methods), AeDES also provides probabilities of exceeding particular thresholds of interest (Figure 5b).

To illustrate the use of both deterministic and probabilistic forecasts, consider the recent Zika epidemic in the Americas³⁰. Official CDC numbers³¹ for Zika cases in the US indicate that both Miami, FL, and New York City (NYC), NY, reported slightly more than 1,000 cases in 2016, around 40% of the total number of cases in the US. Most of these cases were reported after the summer of 2016, a period of increased environmental suitability and human mobility (e.g., tourism to the Caribbean). We will focus on these two cities in the following example.

By the beginning of May 2016, decision makers using AeDES would have expected enhanced suitability conditions for Zika during JJA in basically all of the southeastern US states, but also the Caribbean, most of Central America and northern South America (Figures 4a,b and 5a), where several Zika cases had been already reported. Although it was highly probable that both Miami and NYC exhibited above-normal suitability conditions (Figure 5a), only Miami was expected to exceed $\mathcal{R}_0 = 3$ (Figure 5b). In fact, the decision makers could have used AeDES to determine that most probably Miami would not exceed $\mathcal{R}_0 = 3.4$ (Figure 5c), while NYC most probably would not exceed $\mathcal{R}_0 = 2$ (Figure 4d). These probabilistic forecasts were consistent with the deterministic ones for both cities (Figures 4a,b), and by early September 2016 -once the actual summer $\mathcal{R}_0$ values were available in the monitor sub-system-, the decision makers would have discovered that the forecasts were actually very skillful (Figure 4c). But coming back to May 2016, what those particular $\mathcal{R}_0$ forecasts meant?

Given an original number of 40 Zika cases and a generation time of 20 days (15.6-25.6 days; standard deviation of 7.4 days)³², an $\mathcal{R}_0 = 2$ means that after four generations—each spaced ~ 20 days—, or in about 3 months, there would be a total of 600 local transmission cases related to the original 40. Since $\mathcal{R}_0$ is proportional to the duration of infectivity, an ideal action would be to reduce the infective period of cases, such that the effective reproduction number, $\mathcal{R}$, is reduced. For example, in NYC, with an expected $\mathcal{R}_0 < 2$ for JJA 2016, any combination of strategies to reduce the effective duration of infectivity by over 50% would mean an average $\mathcal{R} < 1$, which should stop the spread of Zika over time. Beyond the obvious vector control strategies (for which knowing in advance when it is not going to rain could be useful), increasing traveler health surveillance, reducing the symptom-onset-to-isolation times, and the mosquito bite rates via specialized clothing and personal protective items can all help decrease the reproduction number. Economic costs for fighting the Zika epidemic would be most probably higher for Miami, given the higher $\mathcal{R}_0$ value forecast for the summer.

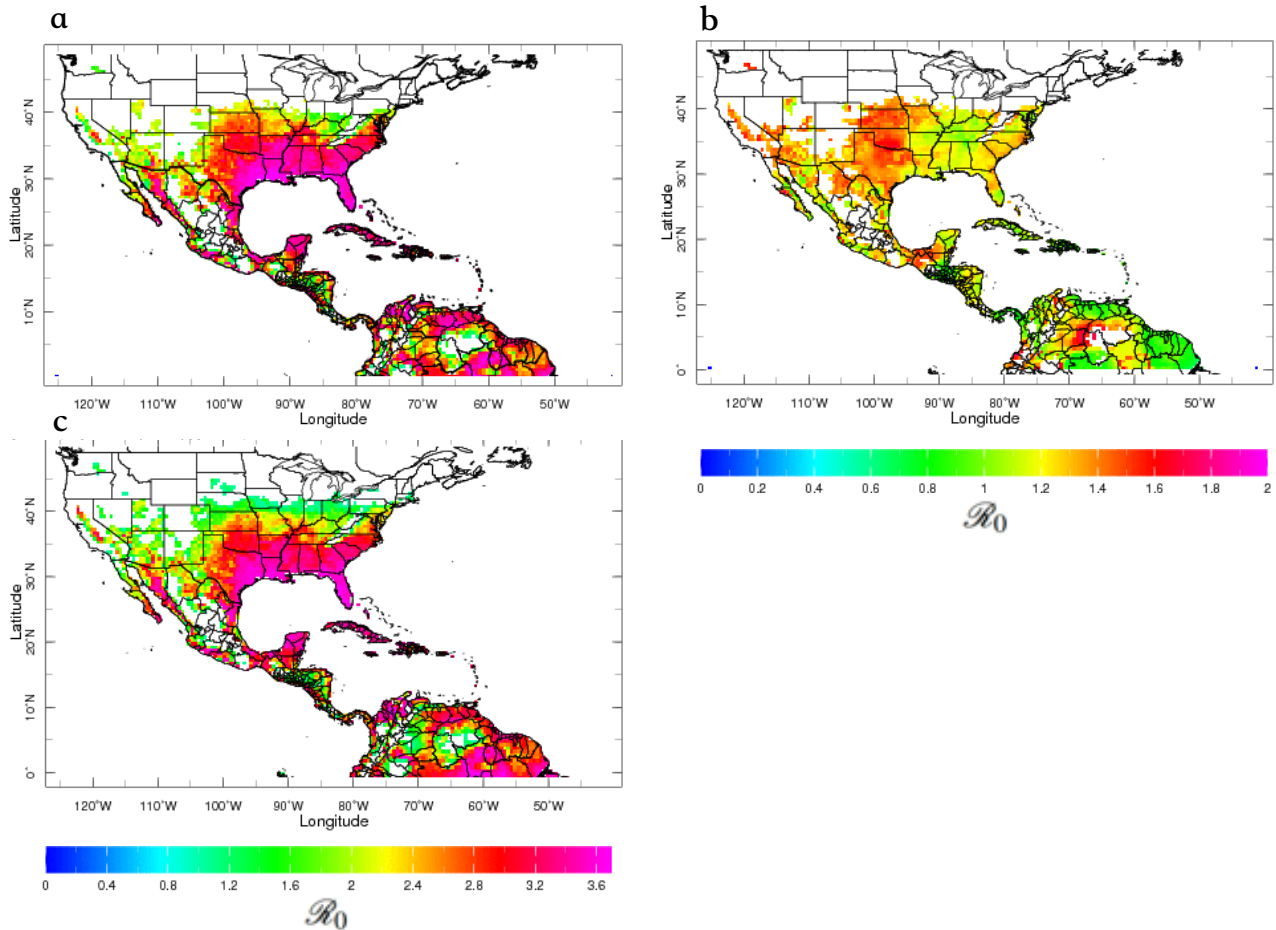


Figure 4. Example of deterministic forecasts for JJA 2016, initialized in May 2016. (a) AeDES seasonal forecast of the expected value of $\mathcal{R}_0$, along with the (b) forecast standard deviation (σ), presented to provide decision-makers information about the forecast uncertainty. (c) Actual seasonal values of $\mathcal{R}_0$, provided by the AeDES monitor sub-system.

A NextGen climate-and-health service for multiple timescales

AeDES is a “next generation” system because (1) it successfully tailors global climate information to be used at regional scales, (2) pattern-based calibration targeting mean, amplitude and spatial biases is performed using a monitoring system based on the same variable that is being predicted, and (3) it produces tailored deterministic and probabilistic forecasts for user-selected thresholds of interest, including the use of the entire probability density function (also known as “forecasts in flexible format”²⁹) to better assess uncertainties.

Previous research^{8,11,16} has underscored the importance of analyzing climate signals at multiple timescales to improve decision-making processes in the health sector. In particular, Muñoz et al¹⁶ and Thomson et al¹¹ have shown that the seasonal-

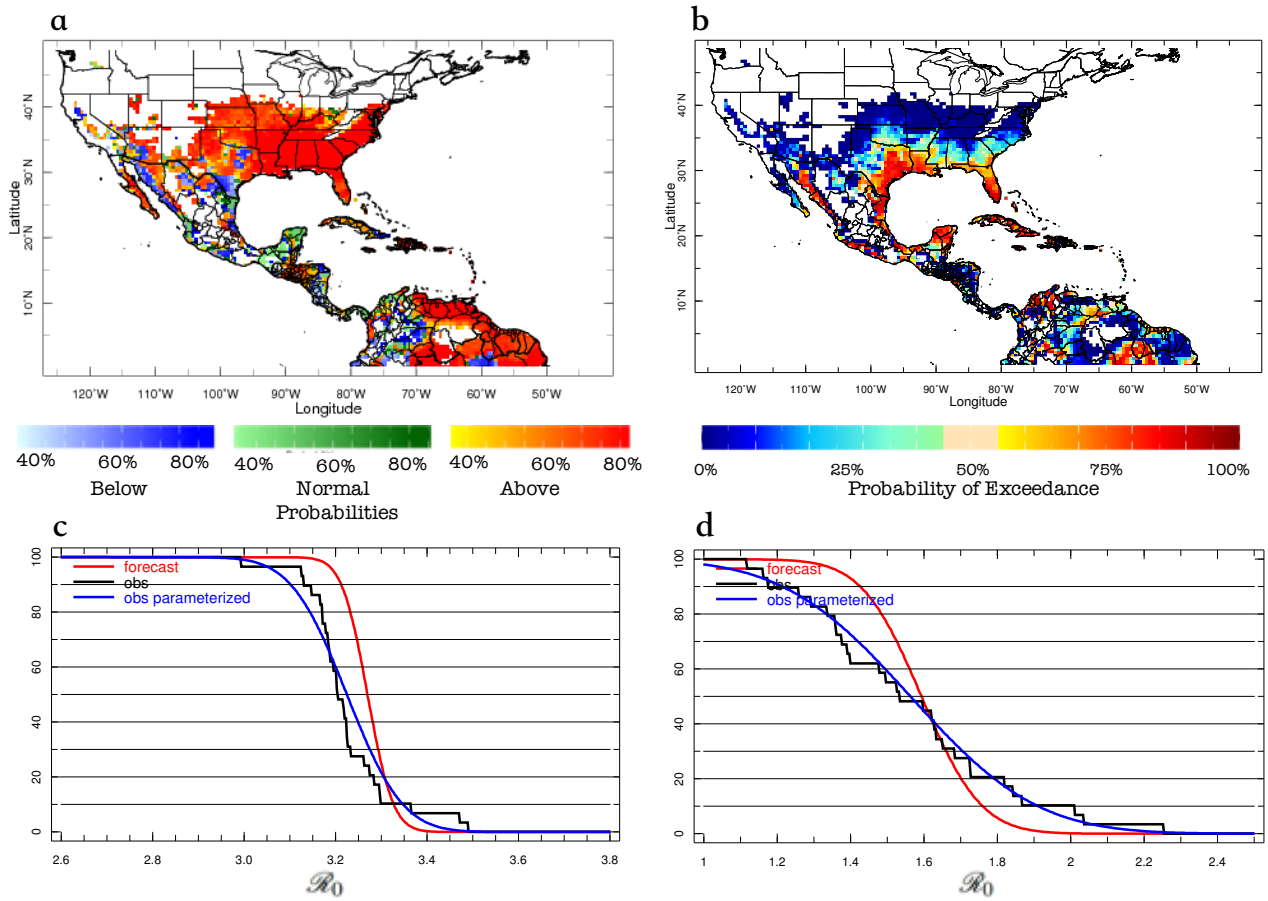


Figure 5. As in Figure 4 but for probabilistic forecasts. (a) Tercile-based forecast; with uncertainty presented in terms of probabilities for each category: below-normal (in red), normal (green) and above-normal (in blue). (b) Spatial forecast of probabilities of exceeding $R_0 = 3$. In the bottom panels, the entire "observed" (blue) and forecast (red) probability of exceedance distributions for (c) Miami, FL, and (d) New York City, NY, are also presented for comparison; the black curves are the empirical "observed" probability of exceedance distributions, before fitting a Gaussian function.

to-interannual timescale tends to explain most of the total variance observed in climate variables impacting vector-borne disease transmission, like temperature and rainfall. Hence, although the long-term climate change and natural decadal variability signals also are considered, AeDES pays special attention to continuously providing actionable information at seasonal-to-interannual timescales, which along with the weather and sub-seasonal³³ scales are the most often used for health early warning systems.

Due to large uncertainties in long-term climate projections, the present approach should in general not be used in combination with climate change scenarios. Nonetheless, the same approach is adequate for shorter-term timescales, like the sub-seasonal (roughly 2-6 weeks³³) or weather (0-2 weeks) ones. Providing actionable information at multiple timescales (e.g., via the IRI's Ready-Set-Go approach²⁹). The team is presently exploring when and where predictive skill at these timescales is high enough to guide decision-making processes in the health sector, taking advantage of windows of opportunities in forecasts at those timescales³⁴.

Methods

Data

All analyses are conducted for the geographical domain defined by the coordinates 126°W-40°W and 1°S-50°N (Figure 1).

Rather than focusing on particular diseases, here we considered common environmental thresholds and ento-epidemiological parameters for *Aedes*-borne diseases as a whole. If the parameters are well known for diseases of interests, then the same approach can be used to have tailored information for those cases. For consistency with previous studies and model validations, we used the same ento-epidemiological parameters reported by Liu-Helmersson et al²³, Wesolowski et al²², Caminade et al.²¹

and Mordecai et al.²⁴. The equations and parameter choice can be found in those references, and in our scripts used to build AeDES (see Data and Codes Availability below).

Two types of near-surface (2 meter) temperature datasets were used: observations and forecasts. Observations consist of gridded fields from NCEP's GHCN-CAMS²⁵ project, at 0.5 degree resolution. When designing the monitoring system, additional observation datasets were used to compare between different spatial resolutions; these products are CRUv4³⁵, at 0.5 degree resolution, and PRISM³⁶, at 0.042 degrees. All products were interpolated to the lower resolution (0.5 degree), and spatial correlations were computed for the 1982-2010 period. No statistically significant differences were observed between the products at $p=0.05$, using a simple t-Student test (not shown).

Given that the GHCN-CAMS dataset is freely updated every month and covers not only North America but also the boundary regions of interest (the Caribbean, Central America and northern South America), this product was selected to force the multi- $\mathcal{R}_0$ model for the monitoring sub-system, in the same way the AeDES hindcasts (described below) were computed, except that only one "realization" (the observed climate) was used. Although $\mathcal{R}_0$ is not a direct observable, for simplicity in this paper we refer to this set of environmental suitability maps as the "observed" (or actual) $\mathcal{R}_0$. The approach is general and does not require GHCN-CAMS to work; if a different reliable observed temperature dataset is available (for example at higher spatial resolution), the system can use it.

The other temperature dataset consists of seasonal predictions from all models operationally available in the North American Multi-Model Ensemble (NMME) project³⁷, as in Muñoz et al.⁸, consisting of a total of 96 ensemble members. AeDES presently uses the latest version of the Canadian climate model (CanSIPv2), after the older Canadian models were discontinued in August 2019.

In addition, the monitoring sub-system also presents infant mortality data from the Socioeconomic Data and Applications Center (SEDAC; <https://sedac.ciesin.columbia.edu/data/collection/povmap>), and the Gridded Population of the World, Version 4 (GPWv4), Revision 11^{38,39}.

$\mathcal{R}_0$ Models and Design of the Next Generation Forecast Sub-System

As indicated, we used four previously-validated $\mathcal{R}_0$ models, described in detail by Liu-Helmersson et al.²³, Wesolowski et al.²², Caminade et al.²¹ and Mordecai et al.²⁴. These models were selected from a set analyzed by Mordecai et al.²⁴, being the ones that better represent the dependence of $\mathcal{R}_0$ with temperature.

To build the $\mathcal{R}_0$ hindcasts, all four $\mathcal{R}_0$ models were forced independently using each one of the 96 NMME climate realizations (i.e., a total of 384 realizations) in hindcast –or retrospective forecast– mode for each season and year in the 1982-2010 period (29 years). For example, all May initializations in the 1982-2010 period were used to obtain the 29 NMME temperature hindcasts JJA seasons, which were then used to force each one of the $\mathcal{R}_0$ models to produce the 1982-2010 environmental suitability hindcasts. Hence, a total of 4,608 (4 $\mathcal{R}_0$ models times 12 initializations times 96 climate model members) 29-yr long seasonal hindcasts were produced.

We then performed a pattern-based calibration for each season and model independently, to avoid mixing models with different characteristics when correcting for mean, amplitude and spatial biases. Following the approach of Muñoz et al.⁸, the calibration method selected was principal component regression⁴⁰. The PCR-based calibration approach builds a regression model for each gridbox of the "observed" $\mathcal{R}_0$ field, using a linear combination of the hindcast $\mathcal{R}_0$'s Empirical Orthogonal Functions (EOFs). The best cross-validated models were identified using a leave-five-out cross-validation window and the Kendall's τ coefficient; these final models are built based on a maximum of 5 EOFs, depending on the season.

The AeDES multi- $\mathcal{R}_0$ calibrated ensemble mean was computed using each calibrated model, providing deterministic outcomes, and the entire probability density function (PDF) for $\mathcal{R}_0$, computing in that case the average of the Gaussian distribution parameters for each grid box (i.e., an ensemble built in the "probability space"). The PDF is used to offer uncertainty information for decision-makers in the health sector and to compute the forecast probability of exceeding thresholds selected by the user. Tailored probabilistic forecasts produced using the entire PDF are often called "forecasts in flexible format"²⁹. These products are available in the AeDES Maproom (<https://aedes.iri.columbia.edu>).

The forecast skill assessment was conducted for each calibrated $\mathcal{R}_0$ model and the AeDES ensemble system for the 1982-2010 period, using the actual $\mathcal{R}_0$ available from monitoring sub-system as reference. Two contrasting seasons were selected for analysis in this study, boreal winter and summer.

The calibration and skill assessment processes were computed using the International Research Institute for Climate and Society's (IRI) Climate Predictability Tool (CPT) version 16.3.2⁴¹, and its Python interface⁴² (PyCPTv1.6; <https://github.com/agmunozs/PyCPT>) to facilitate the mass production of the different hindcasts, skill assessment maps and forecasts. The resulting files were migrated to the IRI Data Library for public archiving and plotting.

Data and Codes Availability

All input and produced data is freely available at the International Research Institute for Climate and Society's Data Library:

<http://iridl.ldeo.columbia.edu/home/.agmunoz/.Aedes/#info>
 Codes are available at Muñoz's GitHub account:
<https://github.com/agmunozs/Vectorbornediseases>

References

1. Cilingiroglu, N. Health, globalization and developing countries. *Saudi medical journal* **26**, 191–200 (2005).
2. Bloom, D. E., Black, S. & Rappuoli, R. Emerging infectious diseases: A proactive approach, DOI: [10.1073/pnas.1701410114](https://doi.org/10.1073/pnas.1701410114) (2017).
3. Kelly-Hope, L. & Thomson, M. C. Climate and infectious diseases. In *Advances in Global Change Research*, vol. 30, 31–70, DOI: [10.1007/978-1-4020-6877-5_3](https://doi.org/10.1007/978-1-4020-6877-5_3) (Springer Netherlands, Dordrecht, 2008).
4. World Health Organization [WHO]. WHO | Global Strategy for dengue prevention and control, 2012–2020. Tech. Rep. (2015).
5. Messina, J. P. *et al.* Mapping global environmental suitability for Zika virus. *eLife* **5**, e15272, DOI: [10.7554/eLife.15272](https://doi.org/10.7554/eLife.15272) (2016).
6. Heesterbeek, H. *et al.* Modeling infectious disease dynamics in the complex landscape of global health, DOI: [10.1126/science.aaa4339](https://doi.org/10.1126/science.aaa4339) (2015).
7. Yang, W., Karspeck, A. & Shaman, J. Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics. *PLoS Comput. Biol.* **10**, e1003583, DOI: [10.1371/journal.pcbi.1003583](https://doi.org/10.1371/journal.pcbi.1003583) (2014).
8. Muñoz, Á. G. *et al.* Could the Recent Zika Epidemic Have Been Predicted? *Front. Microbiol.* **8**, 1291, DOI: [10.3389/fmicb.2017.01291](https://doi.org/10.3389/fmicb.2017.01291) (2017).
9. Lowe, R. *et al.* Nonlinear and delayed impacts of climate on dengue risk in Barbados: A modelling study. *PLoS Medicine* **15**, e1002613, DOI: [10.1371/journal.pmed.1002613](https://doi.org/10.1371/journal.pmed.1002613) (2018).
10. Corley, C. D. *et al.* Disease prediction models and operational readiness. *PLoS ONE* **9**, e91989, DOI: [10.1371/journal.pone.0091989](https://doi.org/10.1371/journal.pone.0091989) (2014).
11. Thomson, M. C., Muñoz, Á. G., Cousin, R. & Shumake-Guillemot, J. Climate drivers of vector-borne diseases in Africa and their relevance to control programmes. *Infect. Dis. Poverty* **7**, 81, DOI: [10.1186/s40249-018-0460-1](https://doi.org/10.1186/s40249-018-0460-1) (2018).
12. Monaghan, A. J. *et al.* On the Seasonal Occurrence and Abundance of the Zika Virus Vector Mosquito *Aedes Aegypti* in the Contiguous United States. *PLoS currents* **8**, 1–31, DOI: [10.1371/currents.outbreaks.50dfc7f46798675fc63e7d7da563da76](https://doi.org/10.1371/currents.outbreaks.50dfc7f46798675fc63e7d7da563da76) (2016).
13. Hales, S., Weinstein, P. & Woodward, A. Dengue fever epidemics in the South Pacific: Driven by El Nino Southern Oscillation? [14], DOI: [10.1016/S0140-6736\(05\)65737-6](https://doi.org/10.1016/S0140-6736(05)65737-6) (1996).
14. Tipayamongkhogul, M., Fang, C. T., Klinchan, S., Liu, C. M. & King, C. C. Effects of the El Niño-Southern Oscillation on dengue epidemics in Thailand, 1996-2005. *BMC Public Heal.* **9**, DOI: [10.1186/1471-2458-9-422](https://doi.org/10.1186/1471-2458-9-422) (2009).
15. Stewart-Ibarra, A. M. & Lowe, R. Climate and non-climate drivers of dengue epidemics in southern coastal Ecuador. *Am. J. Trop. Medicine Hyg.* **88**, 971–981, DOI: [10.4269/ajtmh.12-0478](https://doi.org/10.4269/ajtmh.12-0478) (2013).
16. Muñoz, Á. G., Thomson, M. C., Goddard, L. M. & Aldighieri, S. The Latin American and Caribbean Climate Landscape for ZIKV Transmission. Tech. Rep., International Research Institute for Climate and Society (IRI). Columbia University., New York (2016). DOI: [10.7916/D8X34XHV](https://doi.org/10.7916/D8X34XHV).
17. Muñoz, Á. G., Thomson, M. C., Goddard, L. & Aldighieri, S. Analyzing climate variations at multiple timescales can guide Zika virus response measures. *GigaScience* **5**, 41, DOI: [10.1186/s13742-016-0146-1](https://doi.org/10.1186/s13742-016-0146-1) (2016).
18. WHO. WHO Guidance for Surveillance during an Influenza Pandemic (2018).
19. Chiew, M. *et al.* Estimation du taux de reproduction efficace de la rougeole en Australie à partir de données de notification de routine. *Bull. World Heal. Organ.* **92**, 171–177, DOI: [10.2471/BLT.13.125724](https://doi.org/10.2471/BLT.13.125724) (2014).
20. Dietz, K. The estimation of the basic reproduction number for infectious diseases. *Stat. Methods Med. Res.* **2**, 23–41, DOI: [10.1177/096228029300200103](https://doi.org/10.1177/096228029300200103) (1993).
21. Caminade, C. *et al.* Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proc. Natl. Acad. Sci.* **114**, 119–124, DOI: [10.1073/pnas.1614303114](https://doi.org/10.1073/pnas.1614303114) (2017).

22. Wesolowski, A. *et al.* Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci.* **112**, 11887–11892, DOI: [10.1073/pnas.1504964112](https://doi.org/10.1073/pnas.1504964112) (2015). 329 330
23. Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A. & Rocklöv, J. Vectorial capacity of *Aedes aegypti*: Effects of temperature and implications for global dengue epidemic potential. *PLoS ONE* **9**, e89783, DOI: [10.1371/journal.pone.0089783](https://doi.org/10.1371/journal.pone.0089783) (2014). 331 332 333
24. Mordecai, E. A. *et al.* Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models. *PLoS Neglected Trop. Dis.* **11**, e0005568, DOI: [10.1371/journal.pntd.0005568](https://doi.org/10.1371/journal.pntd.0005568) (2017). 334 335
25. Fan, Y. & van den Dool, H. A global monthly land surface air temperature analysis for 1948-present. *J. Geophys. Res. Atmospheres* **113**, D01103, DOI: [10.1029/2007JD008470](https://doi.org/10.1029/2007JD008470) (2008). 336 337
26. Mason, S. J. & Weigel, A. P. A Generic Forecast Verification Framework for Administrative Purposes. *Mon. Weather. Rev.* **137**, 331–349, DOI: [10.1175/2008MWR2553.1](https://doi.org/10.1175/2008MWR2553.1) (2009). 338 339
27. Monaghan, A. J. *et al.* Consensus and uncertainty in the geographic range of *Aedes aegypti* and *Aedes albopictus* in the contiguous United States: Multi-model assessment and synthesis. *PLoS computational biology* **15**, e1007369, DOI: [10.1371/journal.pcbi.1007369](https://doi.org/10.1371/journal.pcbi.1007369) (2019). 340 341 342
28. van den Driessche, P. Reproduction numbers of infectious disease models. *Infect. Dis. Model.* **2**, 288–303, DOI: [10.1016/j.idm.2017.06.002](https://doi.org/10.1016/j.idm.2017.06.002) (2017). 343 344
29. Goddard, L., Baethgen, W. E., Bhojwani, H. & Robertson, A. W. The International Research Institute for Climate & Society: why, what and how. *Earth Perspectives* **1**, 10, DOI: [10.1186/2194-6434-1-10](https://doi.org/10.1186/2194-6434-1-10) (2014). 345 346
30. Zhang, Q. *et al.* Spread of Zika virus in the Americas. *Proc. Natl. Acad. Sci. United States Am.* **114**, E4334–E4343, DOI: [10.1073/pnas.1620161114](https://doi.org/10.1073/pnas.1620161114) (2017). 347 348
31. CDC. Case Counts in the US | Zika Virus | CDC (2016). 349
32. Ferguson, N. M. *et al.* Countering the Zika epidemic in Latin America, DOI: [10.1126/science.aag0219](https://doi.org/10.1126/science.aag0219) (2016). 350
33. Vitart, F. *et al.* The subseasonal to seasonal (S2S) prediction project database. *Bull. Am. Meteorol. Soc.* **98**, 163–173, DOI: [10.1175/BAMS-D-16-0017.1](https://doi.org/10.1175/BAMS-D-16-0017.1) (2017). 351 352
34. Mariotti, A. *et al.* Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond. *Bull. Am. Meteorol. Soc.* BAMS–D–18–0326.1, DOI: [10.1175/BAMS-D-18-0326.1](https://doi.org/10.1175/BAMS-D-18-0326.1) (2020). 353 354
35. Harris, I., Jones, P., Osborn, T. & Lister, D. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**, 623–642, DOI: [10.1002/joc.3711](https://doi.org/10.1002/joc.3711) (2014). 355 356
36. Di Luzio, M., Johnson, G. L., Daly, C., Eischeid, J. K. & Arnold, J. G. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *J. Appl. Meteorol. Climatol.* **47**, 475–497, DOI: [10.1175/2007JAMC1356.1](https://doi.org/10.1175/2007JAMC1356.1) (2008). 357 358 359
37. Kirtman, B. P. *et al.* The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Am. Meteorol. Soc.* **95**, 585–601, DOI: [10.1175/BAMS-D-12-00050.1](https://doi.org/10.1175/BAMS-D-12-00050.1) (2014). 360 361 362
38. Doxsey-Whitfield, E. *et al.* Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Pap. Appl. Geogr.* **1**, 226–234, DOI: [10.1080/23754931.2015.1014272](https://doi.org/10.1080/23754931.2015.1014272) (2015). 363 364
39. Center for International Earth Science Information Network (CIESIN) Columbia University. Gridded Population of the World, Version 4 (GPWv4): Administrative Unit Center Points with Population Estimates, Revision 11., DOI: [10.7927/H45Q4T5F](https://doi.org/10.7927/H45Q4T5F) (2019). 365 366 367
40. Tippett, M. K., DelSole, T., Mason, S. J. & Barnston, A. G. Regression-based methods for finding coupled patterns. *J. Clim.* **21**, 4384–4398, DOI: [10.1175/2008JCLI2150.1](https://doi.org/10.1175/2008JCLI2150.1) (2008). 368 369
41. Mason, S. J. and Tippett, M. K. and Song, L. and Muñoz, Á. G. Climate Predictability Tool version 16.3.2, DOI: [10.7916/D8-86DY-WQ10](https://doi.org/10.7916/D8-86DY-WQ10) (2020). 370 371
42. Muñoz, Á.G. and Robertson, A.W. and Turkington, T. and Mason, S.J. PyCPT: a Python interface and enhancement for IRI's Climate Predictability Tool, DOI: [10.5281/ZENODO.3551936](https://doi.org/10.5281/ZENODO.3551936) (2019). 372 373

Acknowledgements

The authors acknowledge the NMME project and its data dissemination supported by NOAA, NSF, NASA and DOE, and acknowledge the help of NCEP, IRI and NCAR personnel in creating, updating and maintaining the NMME archive. Á.G.M. was supported by the NOAA grant NA18OAR4310339 and the ARBOPREVENT project (Swedish Research Council Formas grant 2018-01754). X.C. was supported by the NOAA grant NA18OAR4310339. This publication was supported by the Cooperative Agreement Number U01CK000509-01, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NOAA and the Centers for Disease Control and Prevention or the Department of Health and Human Services. The authors are grateful to Dr. Simon Mason, Rémi Cousin and Laurel DiSera for insightful discussions regarding methodological aspects of the study.

Author contributions statement

Á.G.M. designed the monitoring and forecast system, and performed all calculations; X.C. created the public Maproom (web platform) that hosts the monitoring and forecasting sub-systems, and automated the production of operational forecasts; A.R.C. led the decision-making component of the study; P.K. performed part of the literature review and preliminary sensitivity analysis; E.M. contributed with model codes in R language, and analysis; Á.G.M. and M.C.T. wrote the first version of the manuscript; all authors analyzed the results and contributed to the final version.

Additional information

Competing interests The author(s) declare no competing interests.